Diese Website verwendet Cookies von Google, um Dienste anzubieten und den Traffic zu analysieren. Deine IP-Adresse und dein User-Agent werden zusammen mit Messwerten zur Leistung und Sicherheit für Google freigegeben. So können Nutzungsstatistiken generiert, Missbrauchsfälle erkannt und behoben und die Qualität des Dienstes gewährleistet werden.

WEITERE INFORMATIONEN OK

An unending variety of topics

Wednesday, November 01, 2006

Robot law

To what extent can law be mechanized? This author, Mark Miller, Lawrence Lessig, and others have proposed that software code can provide a substitute, to a greater or lesser degree, for various kinds of legal rules, especially in new software-driven areas such as the Internet. Indeed software increasingly provides a *de facto* set of law-like rules governing human interaction even when we don't intend such a result.

This post proposes a framework for mechanical laws in physical spaces inspired by the traditional English common law of trespass (roughly what today is called in common law countries the "intentional torts," especially trespass and battery).

Several decades ago, Isaac Asimov posited three laws to be programmed into robots:

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings, except where such orders would conflict with the First Law, and
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

I reject these as an approach for analyzing the mechanization of law because (i) they represent a law of robot servants towards human masters, which is quite different from my present goal of analyzing the mechanization of laws among equals (whether equal as fellow property or equal as fellow property owner), and (ii) they are hopelessly subjective and computationally intractible. How does a robot determine when a human has been "harmed"? Even if that problem was solved how does a robot predict when a human will come to harm, much less how to avoid their own actions causing such harm, much less what actions will lead to avoidance of that harm? While Asimov attempted to address some of these issues in various oversimplified scenarios in his stories, this is no substitute for the real-world experience of actual disputes and settlements, and the resulting precedents and the evolution of rules.

Thus I propose to substitute for Asimov's Laws a framework which (1) is based on a real-world, highly evolved system of law that stressed concrete and objective rules rather than abstract, ambiguous, and subjective ethical precepts, (2) is "peer-to-peer" rather than "master-servant", and (3) is based, as much as possible, not on hypothetical human-like "artificial intelligence," but on a framework much of which can be tested with hardware and simple algorithms resembling those we possess today. Indeed, a prototype of this system can probably be developed using Lego(tm) robot kits, lasers, photosensors, sonar, video cameras, and other off-the-shelf

Pages

- Home

About Me

Nick Szabo

"A premier thinker about history, law and economics, and the lessons they have for security." -- Adam Shostack, Emergent Chaos

"Szabo comes out with these essays that leave me in awe." -- Brian Dunbar

"Reading material that is eclectic, challenging and endlessly fascinating." -- Sean McGrath, Propylon

"Like most blogs worth my attention, this blog is updated only infrequently. That is because the authors of blogs worth my attention only post when they have something to say that is true, relevant and not already known by their audience. Most of the human race does not have the skill to know when an idea has these three properties. The skill is particularly rare in the fields of politics and economics, which is why this blog is such a rare and valuable thing." -- Richard Hollerith

View my complete profile

hardware along with some custom computer programming using known algorithms.

Using such a prototype framework I hope that we can test out new legal rules and political systems in contests without harming actual humans. I also hope that we can test the extent to which law can be mechanized and to which security technology can prevent or provide strong evidence for breaches of law.

What about possible application to real-world security systems and human laws? In technical terms, I don't expect most of these mechanical laws to be applied in a robotic *per se* fashion to humans, but rather mostly to provide better notice and gather better evidence, perhaps in the clearest cases creating a *prima facie* case shifting the burden of lawsuit and/or the burden of proof to the person against whom clear evidence of law-breaking has been gathered. The system may also be used to prototype "smart treaties" governing weapons systems.

The main anthropomorphic assumption we will make is the idea that our robots can be motivated by incentives, e.g. by the threat of loss of property or of personal damage or destruction. If this seems unrealistic, think of our robots as being agents of humans who after initially creating the robots are off-stage. Under contest rules the humans have an incentive to program their robots to avoid damage to themselves as well as to avoid monetary penalties against their human owners; thus we can effectively treat the robots as being guided (within the limits of flexibility and forethought of simple algorithms) by such incentives.

The traditional trespasses are based on surfaces: a trespass on real property breaks a spatial boundary of that property, and a battery touches the skin or clothes or something closely attached to same (such as a tool grasped in the hand). Corresponding to such laws we have have developed a variety of security devices such as doors, locks, fences, etc. that provide notice of a surface and sometimes provide evidence that such a surface has been forceably crossed. The paradigmatic idea is "tres-pass", Law French for "big step": a crossing of a boundary that, given the notice and affordances provided, the trespasser knew or should have know was a violation of law. In our model, robots should have been programmed to not cross such boundaries without owner consent or other legal authority, ideas to be explored further below.

One way to set up this system is as contests between human programmers deploying their robots into the robot legal system. In a partially automated system there would be some human-refereed "police" and "courts," but an ideal fully-automated version would run without human intervention. Under contest rules in the fully-automated mode, after the human creators, who had an incentive to program the robots to respond to incentives, are done with their programming and the contest has commenced, only robots and other things exist in this robotic legal system. Only robots are legal persons; as such they can own other things and devices as well as spatial regions and things fixed therein (real property).

For each type of tort in my robot legal system, I propose a class of surface or other physical condition and correspondong class(es) of sensors:

(1) battery : physical surface of the robot (sensor detects touch). This touching must be the result of a physical trajectory that was probably not accidental (e.g. bumping a fellow robot on within a crowded space is probably not a battery).

(2) trespass to real property : three-dimensional surface of laser-break sensors bounding a space owned by a robot, or other location-finding sensors that can detect boundary crossings.

(3) trespass or conversion of moveable property: involves touch sensors for trespass to chattels; security against coversion probably also involves using RFID, GPS, or more generally proplets in the moveable property. But this is a harder problem, so call it TBD.

(4) assault is a credible threat of imminent battery : probably involves active and passive viewing devices such as radar, sonar and video and corresponding movement detection and pattern recognition. Specifics TBD (another harder problem).

Blog Archive

- ► 2018 (1)
- ► 2017 (3)
- ► 2016 (4)
- ► 2015 (3)
- ► 2014 (3)
- ► 2013 (3)
- ► 2012 (8)
- ► 2011 (12)
- ► 2010 (9)
- ► 2009 (29)
- ► 2008 (55)
- ► 2007 (47)
- ▼ 2006 (130)
  - ► December (4)
  - ▼ November (7)
    - Are most of us a bit Neanderthal?
    - Are green beards ubiquitous?
    - Underappreciated (ii)
    - Underappreciated
    - Towards smarter democracy
    - A raw idea to avoid being cooked
    - Robot law
  - ► October (15)
  - ► September (15)
  - ► August (20)
  - ► July (5)
  - ► June (10)
  - ► May (4)
  - ► April (3)
  - ► March (11)
  - ► February (10)
  - ► January (26)
- ► 2005 (44)

Note that we have eliminated subjective elements such as "intent" and "harm." The idea is to replace these with deductiions from physical evidence that can be gathered by sensors. For example, evidence of a boundary crossing with force often in itself (*per* se) implies probable intent, and in such cases intent can be dispensed with as a conjunctive element or can be automatically inferred. With well-designed physical triggering conditions and notices and affordances to avoid them, we can avoid subjective elements that would require human judgment and still reach the right verdict in the vast majority of cases. In real world law such notices, affordances, and evidence-gathering would allow us to properly assign the burden of lawsuit to the *prima facie* lawbreaker so that a non-frivolous lawsuit would only occur in a miniscule fraction of cases. Thus security systems adapted to and even simulating real-world legal rules could greatly lower legal costs, but it is to be doubted that they could be entirely automated without at least some residual injustice.

Our robotic contract law is simply software code that the robots generated (or their human programmers have previously coded) and agree to abide by, i.e. smart contracts.

Three basic legal defenses (or justifications) for a surface break, taking of moveable property, or assault are:

(1) consent by the owner, expressed or implied (the former requires only some standard protocol that describes the surface and conditions for activity within that surface, i.e. a smart contract; I'm not sure what the latter category, implied consent, is in robot-land, but the "bumping in a crowded space" is probably an example of implied consent to a touch).

(2) self-defense (which may under certain circumstances include defense of property if the victim was illegally within a surface at the time), or

(3) other legal authorization.

The defense (3) of legal authorization is the most interesting and relates strongly to how robots are incentivized, i.e. punished and how robots are compensated for damages. In other words, what are the "police" and "courts" that enforce the law in this robotic world? Such a punishment is itself an illegal trespass unless it is legally authorized. The defense of legal authorization is also strongly related to the defense (2) of self-defense. That will be the topic, I hope, of a forthcoming post.

Posted by Nick Szabo at 5:23 PM

6 comments:

Joel said...

Built-in radar is a little expensive for keeping track of chattel. This is the next best thing: proximity & rate detection. The author describes both how simple the circuit is to make, and how it allows him to operate his lights by "threatening" them.

http://www.gogglemarks.net/index.php?action=display&tag=fightswitch

11:36 PM

Kenneth Stein said...

Nick, a fundamental issue with your paradigm is that the law is based in hindsight. We reach into the bag of facts of a case and select those facts around which we fashion a narrative.

The narrative serves two purposes. First, it provides for the nexus (causal relation) that's required by some legal test, such as trespass, assault, negligence, etc. Secondly, the narrative provides subjective commentary with regard to the actions that are described, forming an impression in the minds of those tasked with deciding the outcome.

One can shift questions of intent and forseeability onto a responsible human, akin to a products liability argument. The problem is that you're attmepting to take a descriptive system, one in which a narrative is constructed with hindsight to meet specific legal thresholds and you're attempting to use it to prescribe future action.

Law is meaningful because it is a product of human reason and emotion. Most humans possess the socially normative emotions and a measure of empathy, both of which are necessary (in addition to some measure of intellectual capability) to act justly and to judge accordingly.

Artificially separating reason from emotion (as has been the case since the Age of Enlightenment) will move humanity to think and act more like the machines for which we are responsible. It's the only way for us to forsee what such machines might do.

By encoding computers and computer controlled devices with senses that are similar to humans, we create an external represenation of what it is to be a person. Rather than becoming more like the machines, by making machines more like us, humanity provides itself the opportunity to grow.

I have a few ideas on what may happen once we realize the challenges of implementing such a model, and I address some of it at my blog Plexav.com.

Stop by and check out what I have to say in Cooperative Computing and in From Web2.0 to Web2.U

Summum Bonum.

12:44 AM

Nick said...

The introduction of Isaac Asimov's three laws of robotics was a ground breaking event for the imaginations of millions of science-fiction readers. He was merely scattering dreams in fields of simpler thought. Asimov wrote the laws in October 1941. wiki: Runaround

10:56 AM

Graham Asher said...

The word trespass is not 'Law French for "big step"'. It comes from Old French trespas 'passing across'; the first element is derived from Latin 'trans'.

2:16 PM

Graham Asher said...

A brilliant and thought-provoking article.

2:18 PM

Nick Szabo said...

I got my version of the derivation of "trespass" in an admittedly offhand way, by consulting with a French speaker who, I guess, guessed that the Old French came from the Latin "passus" or "step" (actually "two steps") and "tres" was like the modern "tres" an adverbial form of "very." That construction sounds awkward translated literally: "take two steps largely" or similar.

The Latin equivalent phrase for "trespass" in the medieval English literature was "transgressio," literally meaning to move across: compare "agress" (move against), "ingress" (move in), "egress" (move out), "digress" (move apart), "congress" (move together), etc.

I'm somewhat skeptical about Webster's derivation that says "tres" meant "trans" rather than an adverbial version of "very" as in modern French. I can imagine it did mean this to the English lawyers, who readily translated between "trespass" and "transgressio." And I can certainly believe that "passer" was by medieval times as in modern French (to pass) rather than the Latin noun "passus", i.e. "two steps" from which it may have derived. So my revised hunch is that at least the Norman French at the time meant something like "pass substantially" (i.e. as opposed to a _de minimus_ or accidental passage).

Even if my derivation is silly, I think the point I thought it might support is valid: one tries, by looking at evidence such as the direction and magnitude of the movement, and whether there was notice of the surface, to make a distinction between accidential and purposeful crossings of the surface.

12:02 AM

Post a Comment

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.